On the Number of Unbordered Factors

Daniel Goč, Hamoon Mousavi, and Jeffrey Shallit

School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada
{dgoc,hamoon.mousavihaji,shallit}@cs.uwaterloo.ca

Abstract. We illustrate a general technique for enumerating factors of k-automatic sequences by proving a conjecture on the number $f(n)$ of unbordered factors of the Thue-Morse sequence. We show that $f(n) \leq n$ for $n \geq 4$ and that $f(n) = n$ infinitely often. We also give examples of automatic sequences having exactly 2 unbordered factors of every length.

1 Introduction

In this paper, we are concerned with certain factors of k-automatic sequences. Roughly speaking, a sequence $\mathbf{x} = a_0 a_1 a_2 \cdots$ over a finite alphabet Δ is said to be k-automatic if there exists a finite automaton that, on input n expressed in base k, reaches a state with output a_n. Automatic sequences were popularized by a celebrated paper of Cobham [3] and have been widely studied; see [1].

More precisely, let k be an integer ≥ 2, and set $\Sigma_k = \{0, 1, \ldots, k-1\}$. Let $M = (Q, \Sigma_k, \Delta, \delta, q_0, \tau)$ be a deterministic finite automaton with output (DFAO) with transition function $\delta : Q \times \Sigma_k \to Q$ and output function $\tau : Q \to \Delta$. Let $(n)_k$ denote the canonical base-k representation of n, without leading zeros, and starting with the most significant digit. Then we say that M generates the sequence $(a_n)_{n \geq 0}$ if $a_n = \tau(\delta(q_0, (n)_k))$ for all $n \geq 0$.

The prototypical example of a k-automatic sequence is the Thue-Morse sequence $\mathbf{t} = t_0 t_1 t_2 \cdots = 01101001 \cdots$, defined by the relations $t_0 = 0$ and $t_{2n} = t_n$, $t_{2n+1} = 1 - t_n$ for $n \geq 0$. It is generated by the DFAO below in Figure 1.

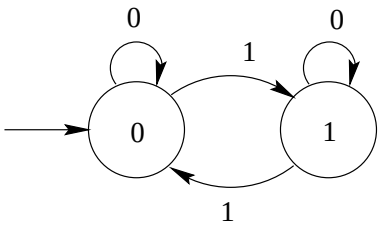

Fig. 1. A finite automaton generating the Thue-Morse sequence $\mathbf{t}$

A *factor* of the sequence $\mathbf{x}$ is a finite word of the form $a_i \cdots a_j$. A finite word w is said to be *bordered* if there is some finite nonempty word $x \neq w$ that is both a prefix and a suffix of w [12,11,6,13]. For example, the English word `ionization` is bordered, as it begins and ends with `ion`. Otherwise w is said to be *unbordered*.

Recently, there has been significant interest in the properties of unbordered factors; see, for example, [9,8,5,10]. In particular, Currie and Saari [4] studied the unbordered factors of the Thue-Morse word.

Currie and Saari [4] proved that if $n \not\equiv 1 \pmod 6$, then the Thue-Morse word has an unbordered factor of length n, but left it open to decide for which lengths congruent to 1 (mod 6) this property holds. This was solved in [7], where the following characterization is given:

Theorem 1. *The Thue-Morse sequence* $\mathbf{t}$ *has an unbordered factor of length* n *if and only if* $(n)_2 \notin 1(01^*0)^*10^*1$.

A harder problem is to come up with an expression for the number of unbordered factors of $\mathbf{t}$. In [2], the second author and co-authors made the following conjecture:

Conjecture 1. Let $f(n)$ denote the number of unbordered factors of length n in $\mathbf{t}$, the Thue-Morse sequence. Then f is given by $f(0) = 1$, $f(1) = 2$, $f(2) = 2$, and the system of recurrences

$$\begin{aligned}
f(4n+1) &= f(2n+1) \\
f(8n+2) &= f(2n+1) - 8f(4n) + f(4n+3) + 4f(8n) \\
f(8n+3) &= 2f(2n) - f(2n+1) + 5f(4n) + f(4n+2) - 3f(8n) \\
f(8n+4) &= -4f(4n) + 2f(4n+2) + 2f(8n) \\
f(8n+6) &= 2f(2n) - f(2n+1) + f(4n) + f(4n+2) + f(4n+3) - f(8n) \\
f(16n) &= -2f(4n) + 3f(8n) \qquad (1) \\
f(16n+7) &= -2f(2n) + f(2n+1) - 5f(4n) + f(4n+2) + 3f(8n) \\
f(16n+8) &= -8f(4n) + 4f(4n+2) + 4f(8n) \\
f(16n+15) &= -8f(4n) + 2f(4n+3) + 4f(8n) + f(8n+7).
\end{aligned}$$

for $n \geq 0$.

This conjecture was obtained by computing a large number of values of f and then looking for possible linear relations among subsequences of the form $(f(2^i n + j))_{n \geq 0}$.

This system suffices to calculate f efficiently, in $O(\log n)$ arithmetic steps.

We now summarize the rest of the paper. In Section 2, we prove Conjecture 1. In Section 3, we discuss how to obtain relations like those above for a given k-regular sequence. In Section 4 we discuss the growth rate of f in detail. Finally, in Section 5, we give examples of other sequences with interesting numbers of unbordered factors.

2 Proof of the conjecture

We now outline our computational proof of Conjecture 1.

First, we need a little notation. We extend the notion of canonical base-k representation of a single non-negative integer to tuples of such integers. For example, by $(m, n)_k$ we mean the unique word over the alphabet $\Sigma_k \times \Sigma_k$ such that the projection π_1 onto the first coordinate gives the base-k representation of m, and the projection π_2 onto the second co-ordinate gives the base-k representation of n, where the shorter representation is padded with leading 0's, if necessary, so that the representations have the same length. For example, $(43, 17)_2 = [1, 0][0, 1][1, 0][0, 1][1, 0][1, 1]$.

Proof. Step 1: Using the ideas in [7], we created an automaton A of 23 states that accepts the language L of all words $(n, i)_2$ such that there is a "novel" unbordered factor of length n in $\mathbf{t}$ beginning at position i. Here "novel" means that this factor does not previously appear in any position to the left. Thus, the number of such words with first component equal to $(n)_2$ equals $f(n)$, the number of unbordered factors of $\mathbf{t}$ of length n. This automaton is illustrated below in Figure 2 (rotated to fit the figure more clearly).

Step 2: Using the ideas in [2], we now know that f is a 2-regular sequence, with a "linear representation" that can be deduced from the structure of A. This gives matrices M_0, M_1 of dimension 23 and vectors v, w such that $f(n) = vM_{a_1} \cdots M_{a_i} w$ where $a_1 \cdots a_i$ is the base-2 representation of n, written with the most significant digit first. They are given below.

$$M_0 = \begin{bmatrix}
1&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&1&0&0\\
0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0
\end{bmatrix}$$

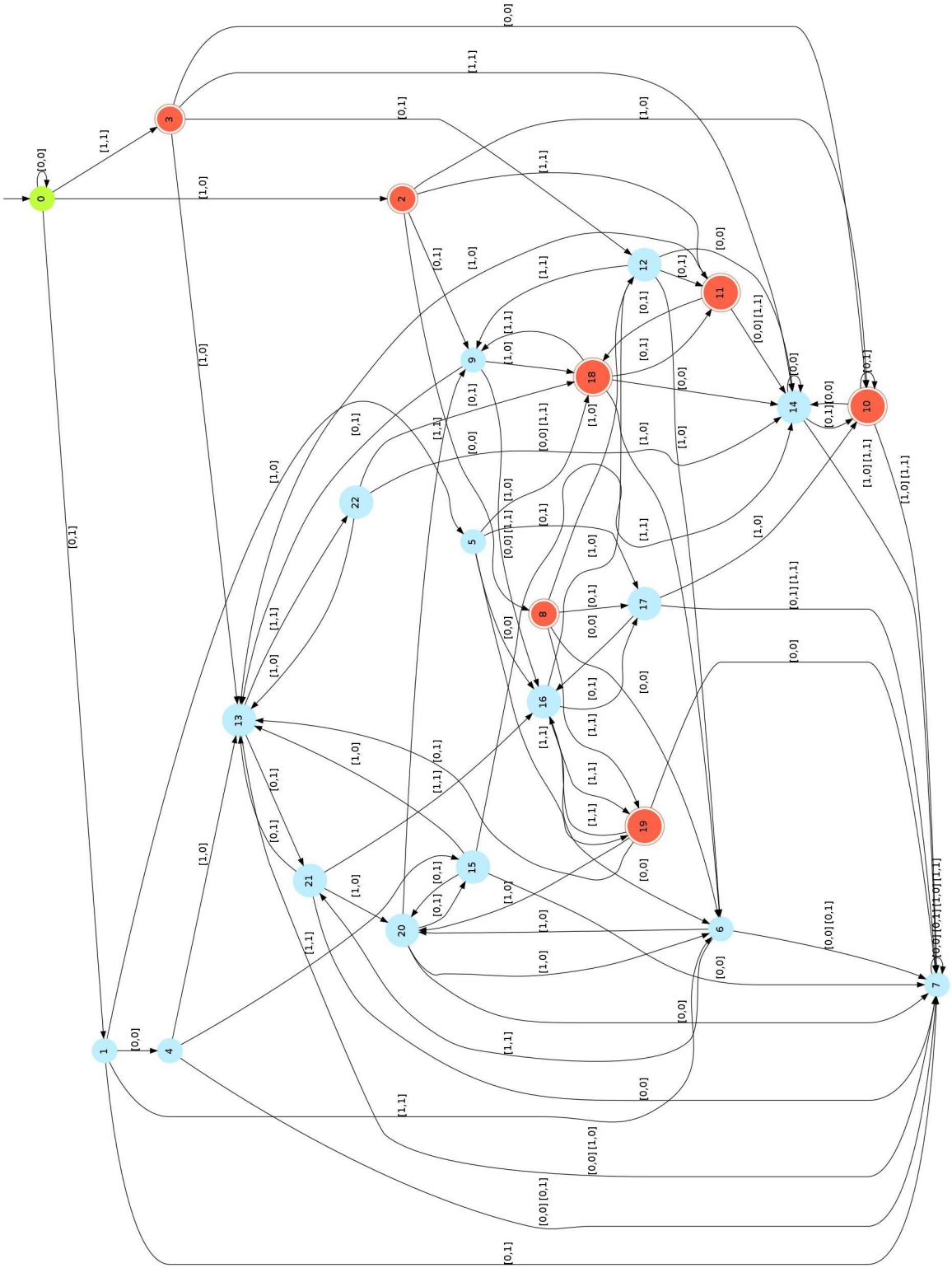

Fig. 2. Automaton accepting $(n, i)_2$ such that there is a novel unbordered factor of length n at position i of $\mathbf{t}$

$$M_1 = \begin{bmatrix}
0&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{bmatrix}.$$

$$v = [1,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0]$$
$$w = [0,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]^T$$

Step 3: Now each of the identities in (1) corresponds to a certain identity in matrices. For example, the identity $f(16n) = -2f(4n) + 3f(8n)$ can be written as

$$vMM_0M_0M_0M_0w = -2vMM_0M_0w + 3vMM_0M_0M_0w, \tag{2}$$

where M is the matrix product corresponding to the base-2 expansion of n. More generally, we can think of M as some arbitrary product of the matrices M_0 and M_1, starting with at least one M_1; this corresponds to an arbitrary $n \geq 1$. We can think of M as a matrix of indeterminates. Then (2) represents an assertion about the entries of M which can be verified. Of course, the entries of M are not completely arbitrary, since they come about as M_1 times some product of M_0 and M_1. We can compute the (positive) transitive closure of $M_0 + M_1$ and then multiply on the left by M_1; the entries that have 0's will be 0 in any product of M_1 times a product of the matrices M_0 and M_1. Thus we can replace the corresponding indeterminates by 0, which makes verifying (2) easier.

Another approach, which is even simpler, is to consider vM in place of M. This reduces the number of entries it is required to check from d^2 to d, where d is the dimension of the matrices.

Step 4: Finally, we have to verify the identies for $n = 0$ and $n = 1$, which is easy.

We carried out this computation in Maple for the matrices M_0 and M_1 corresponding to A, which completes the proof. The Maple program can be downloaded from

`http://www.cs.uwaterloo.ca/~shallit/papers.html` .

3 Determining the relations

The verification method of the previous section can be extended to a method to mechanically *find* the relations for *any* given k-regular sequence g (instead of guessing them and verifying them), given the linear representation of g.

Suppose we are given the linear representation of a k-regular sequence g, that is, vectors v, w and matrices $M_0, M_1, \ldots, M_{k-1}$ such that $g(n) = vM_{a_1}M_{a_2}\cdots M_{a_j}w$, where $a_1a_2\cdots a_j = (n)_k$.

Now let M be arbitrary and consider vM as a vector with variable entries, say $[a_1, a_2, \ldots, a_d]$. Successively compute vMM_yw for words y of length $0, 1, 2, \ldots$ over $\Sigma_k = \{0, 1, \ldots, k-1\}$; this will give an expression in terms of the variables $a_1, \ldots, a_d$. After at most $d+1$ such relations, we find an expression for vMM_yw for some y as a linear combination of previously computed expressions. When this happens, you no longer need to consider any expression having y as a suffix. Eventually the procedure halts, and this corresponds to a system of equations like that in (1).

Consider the following example. Let $k = 2$, $v = [6, 1]$, $w = [2, 4]^T$, and

$$M_0 = \begin{bmatrix} -3 & 1 \\ 1 & 4 \end{bmatrix}$$
$$M_1 = \begin{bmatrix} 0 & 2 \\ -3 & 1 \end{bmatrix}$$

Suppose M is some product of M_0 and M_1, and suppose $vM = [a, b]$.
We find

$$\begin{aligned} vMw &= 2a + 4b \\ vMM_0w &= -2a + 18b \\ vMM_1w &= -8a - 2b \\ vMM_0M_0w &= 24a + 70b \\ vMM_1M_0w &= 36a + 24b \end{aligned}$$

and, solving the linear systems, we get

$$\begin{aligned} vMM_1w &= \frac{35}{11}vMw - \frac{9}{11}vM_0w \\ vMM_0M_0w &= 13vMw + vM_0w \\ vMM_1M_0w &= \frac{174}{11}vMw - \frac{24}{11}vM_0w. \end{aligned}$$

This gives us

$$\begin{aligned} g(2n+1) &= \frac{35}{11}g(n) + \frac{9}{11}g(2n) \\ g(4n) &= 13g(n) + g(2n) \\ g(4n+2) &= \frac{174}{11}g(n) - \frac{24}{11}g(2n) \end{aligned}$$

for $n \geq 1$.

4 The growth rate of $f(n)$

We now return to $f(n)$, the number of unbordered factors of $\mathbf{t}$ of length n. Here is a brief table of $f(n)$:

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
$f(n)$	1	2	2	4	2	4	6	0	4	4	4	4	12	0	4	4	8	4	8	0	8	4	4	8	24	0	4	4	8	4	8	4

Kalle Saari (personal communication) asked about the growth rate of $f(n)$. The following results characterizes it.

Theorem 2. *We have $f(n) \leq n$ for $n \geq 4$. Furthermore, $f(n) = n$ infinitely often. Thus,* $\limsup_{n\geq 1} f(n)/n = 1$.

Proof. We start by verifying the following relations:

$$f(4n) = 2f(2n), \qquad (n \geq 2) \tag{3}$$
$$f(4n+1) = f(2n+1), \qquad (n \geq 0) \tag{4}$$
$$f(8n+2) = f(2n+1) + f(4n+3), \qquad (n \geq 1) \tag{5}$$
$$f(8n+3) = -f(2n+1) + f(4n+2) \qquad (n \geq 2) \tag{6}$$
$$f(8n+6) = -f(2n+1) + f(4n+2) + f(4n+3) \qquad (n \geq 2) \tag{7}$$
$$f(8n+7) = 2f(2n+1) + f(4n+3) \qquad (n \geq 3) \tag{8}$$

These can be verified in exactly the same way that we verified the system (1) earlier.

We now verify, by induction on n, that $f(n) \leq n$ for $n \geq 4$. The base case is $n = 4$, and $f(4) = 2$. Now assume $n \geq 5$. Otherwise,

- If $n \equiv 0 \pmod 4$, say $n = 4m$ and $m \geq 2$. Then $f(4m) = 2f(2m) \leq 2 \cdot 2m \leq 4m$ by (3) and induction.
- If $n \equiv 1 \pmod 4$, say $n = 4m + 1$ for $m \geq 1$, then $f(4m+1) = f(2m+1)$ by (4). But $f(2m+1) \leq 2m+1$ by induction for $m \geq 2$. The case $m = 1$ corresponds to $f(5) = 4 \leq 5$.
- If $n \equiv 2 \pmod 8$, say $n = 8m + 2$, then for $m \geq 2$ we have $f(8m+2) = f(2m+1) + f(4m+3) \leq 6m + 4$ by induction, which is less than $8m + 2$. If $m = 1$, then $f(10) = 4 < 10$.
- If $n \equiv 3 \pmod 8$, say $n = 8m + 3$ for $m \geq 1$, then $f(8m+3) = -f(2m+1) + f(4m+2) \leq f(4m+2) \leq 4m + 2$ by induction.
- If $n \equiv 6 \pmod 8$, say $n = 8m + 6$, then $f(8m+6) = -f(2m+1) + f(4m+2) + f(4m+3) \leq f(4m+2) + f(4m+3) \leq 8m + 5$ by induction, provided $m \geq 2$. For $m = 0$ we have $f(6) = 6$ and for $m = 1$ we have $f(14) = 4$.

– If $n \equiv 7 \pmod 8$, say $n = 8m+7$, then $f(8m+7) = 2f(2m+1)+f(4m+3) \leq 2(2m+1)+4m+3 = 8m+5$ for $m \geq 3$, by induction. The cases $m = 0, 1, 2$ can be verified by inspection.

This completes the proof that $f(n) \leq n$.

It remains to see that $f(n) = n$ infinitely often. We do this by showing that $f(n) = n$ for n of the form $3 \cdot 2^i$, $i \geq 1$. Let us prove this by induction on i. It is true for $i = 1$ since $f(6) = 6$. Otherwise $i \geq 2$, and using (3) we have $f(3 \cdot 2^{i+1}) = 2f(3 \cdot 2^i) = 2 \cdot 3 \cdot 2^i = 3 \cdot 2^{i+1}$ by induction. This also implies the claim $\limsup_{n \geq 1} f(n)/n = 1$.

5 Unbordered factors of other sequences

We can carry out similar computations for other famous sequences. In some cases the automata and the corresponding matrices are very large, which renders the computations time-consuming and the asymptotic behavior less transparent. We report on some of these computations, omitting the details.

Theorem 3. *Let* $\mathbf{r} = r_0 r_1 r_2 \cdots = 00010010 \cdots$ *denote the Rudin-Shapiro sequence, defined by* $r_n =$ *the number of occurrences, taken modulo* 2*, of '11' in the binary expansion of* n*. Let* $f_{\mathbf{r}}(n)$ *denote the number of unbordered factors of length* n *in* $\mathbf{r}$*. Then* $f_{\mathbf{r}}(n) \leq \frac{21}{8} n$ *for all* $n \geq 1$*. Furthermore if* $n = 2^i + 1$*, then* $f(n) = 21 \cdot 2^{i-3}$ *for* $i \geq 4$*.*

Theorem 4. *Let* $\mathbf{p} = p_0 p_1 p_2 \cdots = 0100 \cdots$ *be the so-called "period-doubling" sequence, defined by*

$$p_n = \begin{cases} 1, & \text{if } t_n = t_{n+1}; \\ 0, & \text{otherwise,} \end{cases}$$

where $t_0 t_1 t_2 \cdots$ *is the Thue-Morse word* $\mathbf{t}$*. Note that* $\mathbf{p}$ *is the fixed point of the morphism* $0 \to 01$ *and* $1 \to 00$*. Then* $f_{\mathbf{p}}(n)$*, the number of unbordered factors of* $\mathbf{p}$ *of length* n*, is equal to* 2 *for all* $n \geq 1$*.*

The period-doubling sequence can be generalized to base $k \geq 2$, as follows:

$$\mathbf{p}_k := (\nu_k(n+1) \bmod 2)_{n \geq 0},$$

where $\nu_k(x)$ is the exponent of the largest power of k dividing x. For each k, the corresponding sequence $\mathbf{p}_k$ is a binary sequence that is k-automatic:

Theorem 5. *Let* k *be an integer* ≥ 2*. The sequence* $\mathbf{p}_k$ *is the fixed point of the morphism* φ_k*, where*

$$\varphi_k(0) = 0^{k-1}\,1$$
$$\varphi_k(1) = 0^k.$$

Proof. Note that $\mathbf{p}_k(n) = c$ iff $\nu_k(n+1) = 2j + c$ for some integer $j \geq 0$, and $c \in \{0, 1\}$.

If $0 \leq a < k-1$, then $\mathbf{p}_k(kn+a) = \nu_k(kn+a+1) \bmod 2 = 0$. If $a = k-1$ we have $\mathbf{p}_k(kn+a) = \nu_k(kn+k) \bmod 2 = \nu_k(k(n+1)) \bmod 2 = (2j+c+1) \bmod 2$. Hence if $\mathbf{p}_k(n) = 0$, then $\mathbf{p}_k[kn..kn+k-1] = 0^{k-1}\,1$, while if $\mathbf{p}_k(n) = 1$, then $\mathbf{p}_k[kn..kn+k-1] = 0^k$. It follows that $\mathbf{p}_k$ is the fixed point of φ_k.

The generalized sequence $\mathbf{p}_k$ has the same property of unbordered factors as the period-doubling sequence:

Theorem 6. *The number of unbordered factors of $\mathbf{p}_k$ of length n, for $k \geq 2$ and $n \geq 1$, is equal to 2, and the two unbordered factors are reversals of each other.*

We begin with some useful lemmas.

Lemma 1. *Let $x \in \{0,1\}^*$ be a word. Then $0^{k-1}\,\varphi_k(x)^R = \varphi_k(x^R)\,0^{k-1}$.*

Proof. Suppose $x = a_1 a_2 \cdots a_n$, where each $a_i \in \{0,1\}$. If $a \in \{0,1\}$, let $\overline{a}$ denote $1-a$. Then

$$\begin{aligned}
0^{k-1}\,\varphi_k(x)^R &= 0^{k-1} \left(\prod_{1 \leq i \leq n} \varphi_k(a_i) \right)^R \\
&= 0^{k-1} \left(\prod_{1 \leq i \leq n} 0^{k-1}\,\overline{a_i} \right)^R \\
&= 0^{k-1} \left(\prod_{1 \leq i \leq n} \overline{a_{n+1-i}}\,0^{k-1} \right) \\
&= \left(\prod_{1 \leq i \leq n} 0^{k-1}\,\overline{a_{n+1-i}} \right) 0^{k-1} \\
&= \left(\prod_{1 \leq i \leq n} \varphi_k(a_{n+1-i}) \right) 0^{k-1} \\
&= \varphi_k(x^R)\,0^{k-1}.
\end{aligned}$$

Lemma 2. *If the word w is bordered, then $\varphi_k(w)$ is bordered.*

Proof. If w is bordered, then $w = xyx$ for $x \neq \epsilon$. Then $\varphi_k(w) = \varphi_k(x)\varphi_k(y)\varphi_k(x)$ is bordered.

Lemma 3. *If w is a factor of $\mathbf{p}_k$, then so is w^R.*

Proof. If w is a factor of $\mathbf{p}_k$, then it is a factor of some prefix $\mathbf{p}_k[0..k^i-1]$ for some $i \geq 1$. So it suffices to show that $\mathbf{p}_k[0..k^i-1]^R$ appears as a factor of $\mathbf{p}_k$. In fact, we claim that

$$\mathbf{p}_k[0..k^i-1]^R = \mathbf{p}_k[k^i-1..2k^i-2].$$

To see this, it suffices to observe that $\nu_k(k^i-a) = \nu_k(k^i+a)$ for $0 \leq a < k^i$.

The following lemma describes the unbordered factors of φ_k. If $w = 0^a x$, then by $0^{-a}\, w$ we mean the word x.

Lemma 4. *(a) If w is an unbordered factor of $\mathbf{p}_k$ and $|w| \equiv 0 \pmod{k}$, then $w = \varphi_k(x)$ or $w = \varphi_k(x)^R$, for some unbordered factor x of $\mathbf{p}_k$ with $|x| = |w|/k$.*

(b) If w is an unbordered factor of $\mathbf{p}_k$ and $|w| \equiv a \pmod{k}$ for $0 < a < k$, then $w = 0^{a-k}\,\varphi_k(x)$ or $w = \varphi_k(x)^R\, 0^{a-k}$, for some unbordered factor x of $\mathbf{p}_k$ with $|x| = (|w| - a)/k + 1$.

Proof. (a): Suppose that $w = \mathbf{p}_k[i..i + kn - 1]$ for some integer i. There are two cases to consider: $\mathbf{p}_k[i] = 0$ and $\mathbf{p}_k[i] = 1$.

Suppose $\mathbf{p}_k[i] = 0$. Since $\mathbf{w}$ is unbordered, we have $\mathbf{p}_k[i + kn - 1] = 1$. Then $\nu_k(i + kn) \geq 1$, so $i + kn = km$ for some $m \geq 0$. Then $i = k(m - n)$ is a multiple of k, so $w = \varphi_k(x)$, where $x = \mathbf{p}_k[i/k..i/k+n-1]$. Note that $|x| = |w|/k$. Finally, Lemma 2 shows that x is unbordered.

Suppose $\mathbf{p}_k[i] = 1$. Since w is unbordered, we have $\mathbf{p}_k[i + kn - 1] = 0$. From Lemma 3 we know that w^R is also a factor of $\mathbf{p}_k$ (and also is unbordered). Then from the previous paragraph, we see that $w^R = \varphi_k(x)$ for some unbordered factor x of $\mathbf{p}_k$, with $|x| = |w|/k$. Then $w = \varphi_k(x)^R$, as desired.

(b): Suppose that $w = \mathbf{p}_k[i..i + kn + a - 1]$ for $0 < a < k$. There are two cases to consider: $\mathbf{p}_k[i] = 0$ and $\mathbf{p}_k[i] = 1$.

Suppose that $\mathbf{p}_k[i] = 0$. Since w is unbordered, we know that $\mathbf{p}_k[i + kn + a - 1] = 1$. Then $\nu_k(i + kn + a) \geq 1$, so $i + kn + a = km$ for some $m \geq 0$. Then $i - (k - a) = k(m - n - 1)$ is a multiple of k. Hence

$$0^{k-a}\, w = \mathbf{p}_k[i - (k-a)..i + kn + a - 1] = \varphi_k(\mathbf{p}_k[(i+a)/k - 1..(i+a)/k + n - 1]).$$

Let $x = \mathbf{p}_k[(i + a)/k - 1..(i + a)/k + n - 1]$. Then $w = 0^{a-k}\,\varphi_k(x)$, and $|x| = (|w| - a)/k + 1$. If x is bordered, then using Lemma 2 we have that $0^{k-a}\, w$ has a border of length $\geq k$, so w has a border of length at least a, a contradiction.

Suppose that $\mathbf{p}_k[i] = 1$. Since w is unbordered, we know that $\mathbf{p}_k[i + kn + a - 1] = 0$. Then by Lemma 3 we know that w^R is also an unbordered factor of $\mathbf{p}_k$. Then from the previous paragraph, we get that $w^R = 0^{a-k}\,\varphi_k(x)$ for some unbordered factor x of $\mathbf{p}_k$ where $|x| = (|w| - a)/k + 1$. So $w = \varphi_k(x)^R\, 0^{a-k}$, as desired.

Lemma 5. *Let x be a word and $w = 1x0$ be an unbordered word. Then $0^i\varphi_k(x0)$ is unbordered for $1 \leq i \leq k$.*

Proof. If $i = k$ then $0^k\varphi_k(x0) = \varphi_k(1x0) = \varphi_k(w)$. Suppose $\varphi_k(w)$ is bordered; then there exist $u \neq \epsilon$ and v such that $\varphi_k(w) = uvu$. Since $\varphi_k(0) = 0^{k-1}1$, we know u ends in 1. But since u is a prefix of $\varphi_k(w)$ that ends in 1, it follows that $|u| \equiv 0 \pmod{k}$, and so u is the image of some word r under φ_k. Hence w begins and ends with r, a contradiction.

Now assume $1 \leq i < k$ and $0^i\varphi_k(x0)$ is bordered. Then there exist $u \neq \epsilon$ and v such that $0^i\varphi_k(x0) = uvu$; note that u must end in 1. It follows that

$$\varphi_k(w) = \varphi_k(1x0) = 0^k\varphi_k(x0) = 0^{k-i}(0^i\varphi_k(x0)) = 0^{k-i}uvu.$$

Since $0^{k-i}u$ and $0^{k-i}uvu$ both end in 1 and $0^{k-i}uvu = \varphi_k(w)$, we have $|vu| \equiv 0 \pmod{k}$. Hence $|u| \equiv i \pmod{k}$. It follows that $0^{k-i}uv$ ends in 0^k, so $0^{k-i}uvu = \varphi_k(w)$ begins and ends in $0^{k-i}u$, a contradiction.

We are now ready for the proof of Theorem 6.

Proof. First, we show that there is at least one unbordered factor of every length, by induction on n. The base cases are $n < 2k$, and are left to the reader. Otherwise $n \geq 2k$. Write $n = kn' + i$ where $1 \leq i \leq k$. By induction there is an unbordered word w of length $n' + 1$. Using Lemma 3, we can assume that w begins with 1 and ends with 0, say $w = 1x0$. By Lemma 5 we have that $0^i\varphi_k(x0)$ is unbordered, and it is of length $i + kn' = n$.

It remains to prove there are exactly 2 unbordered factors of every length.

If $n \leq 2k$, then it is easy to see that the only unbordered factors are $1\,0^{n-1}$ and $0^{n-1}\,1$.

Now assume $n > 2k$ and that there are only two unbordered factors of length n' for all $n' < n$; we prove it for n. Let w be an unbordered factor of length n.

If $n \equiv 0 \pmod{k}$, then by Lemma 4 (a), we know that either $w = \phi_k(x)$ or $w = \phi_k(x)^R$, where x is an unbordered factor of length n/k. By induction there are exactly 2 unbordered factors of length n/k; by Lemma 3 they are reverses of each other. Let x be such an unbordered factor; since $|x| = n/k > 2$, either x begins with 0 and ends with 1, or begins with 1 and ends with 0. In the former case, the image $w = \varphi_k(x)$ begins and ends with 0, a contradiction. So x begins with 1 and ends with 0. But there is only one such factor, so there are only two possibilities for w.

Otherwise let $a = n \bmod k$; then $0 < a < k$. By Lemma 4 (b), we know that $w = 0^{a-k}\,\varphi_k(x)$ or $w = \varphi_k(x)^R\,0^{a-k}$, where x is an unbordered factor of length $(|w| - a)/k + 1 \geq 2$. By induction there are exactly 2 such unbordered words; by Lemma 3 they are reverses of each other. Let x be such an unbordered factor; then either x begins with 0 and ends with 1, or begins with 1 and ends with 0. Let us call them x_0 and x_1, respectively, with $x_0 = x_1^R$. Now $\varphi_k(x_0)$ begins with $0^{k-1}\,1$, and ends with 0^k. Hence, provided $a \neq 1$, we see that $w = 0^{a-k}\,\varphi_k(x_0)$ begins with 0 and ends with 0, a contradiction. If $a = 1$, Lemma 4 (b) gives the two factors $0^{1-k}\,\varphi_k(x_0)$ and $\varphi_k(x_0)^R\,0^{1-k}$. The former begins with 1 and ends with 0; the latter begins with 0 and ends with 1.

In the latter case, x_1 begins with 1 and ends with 0. There is only one such x_1 (by induction), and then either $w = 0^{a-k}\,\varphi_k(x_1)$ or $w = \varphi_k(x_1)^R\,0^{a-k}$, giving at most two possibilities for w. In the case $a = 1$, these two factors would seem to give a total of four factors of length n. However, there are only two, since

$$0^{1-k}\,\varphi_k(x_0) = 0^{1-k}\,\varphi_k(x_1^R) = \varphi_k(x_1)^R\,0^{1-k}$$
$$\varphi_k(x_0)^R\,0^{1-k} = 0^{1-k}\,\varphi_k(x_0^R) = 0^{1-k}\,\varphi_k(x_1)$$

This completes the proof.

References

1. J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, 2003.
2. E. Charlier, N. Rampersad, and J. Shallit. Enumeration and decidable properties of automatic sequences. In *DLT 2011*, Vol. 6795 of *Lecture Notes in Computer Science*, pp. 165–179. Springer-Verlag, 2011.
3. A. Cobham. Uniform tag sequences. *Math. Systems Theory* **6** (1972), 164–192.
4. J. D. Currie and K. Saari. Least periods of factors of infinite words. *RAIRO Inform. Théor. App.* **43**(1) (2009), 165–178.
5. J.-P. Duval, T. Harju, and D. Nowotka. Unbordered factors and Lyndon words. *Discrete Math.* **308** (2008), 2261–2264.
6. A. Ehrenfeucht and D. M. Silberger. Periodicity and unbordered segments of words. *Discrete Math.* **26** (1979), 101–109.
7. D. Goč, D. Henshall, and J. Shallit. Automatic theorem-proving in combinatorics on words. In *CIAA 2012*, Vol. 7381 of *Lecture Notes in Computer Science*, pp. 180–191. Springer-Verlag, 2012.
8. T. Harju and D. Nowotka. Periodicity and unbordered words: a proof of the extended Duval conjecture. *J. Assoc. Comput. Mach.* **54** (2007), Article 20.
9. S. Holub. A proof of the extended Duval's conjecture. *Theoret. Comput. Sci.* **339** (2005), 61–67.
10. S. Holub and D. Nowotka. On the relation between periodicity and unbordered factors of finite words. *Internat. J. Found. Comp. Sci.* **21** (2010), 633–645.
11. P. T. Nielsen. A note on bifix-free sequences. *IEEE Trans. Inform. Theory* **IT-19** (1973), 704–706.
12. D. M. Silberger. Borders and roots of a word. *Portugal. Math.* **30** (1971), 191–199.
13. D. M. Silberger. How many unbordered words? *Ann. Soc. Math. Polon. Ser. I: Comment. Math.* **22** (1980), 143–145.